DATA AND METHODS

DATA AND METHODS

This report is based on data from the 1980, 1990, and 2000 decennial censuses. The main methodological issues involved in analyzing racial and ethnic residential segregation revolve around the definition of racial and ethnic categories, geographic boundaries, and segregation measures. We begin with a discussion of these issues, and then follow with a more detailed description of the data and notes on statistical testing and the interpretation of findings.

A. RACE AND ETHNICITY MEASUREMENT

One issue that arises when measuring residential segregation is choosing a reference group against which the segregation of other groups can be measured. We chose non-Hispanic Whites as the reference group (Massey and Denton 1988). For 2000 data, when individuals could report more than one race, we chose those who designated White alone as their racial classification, and not Hispanic.

For other groups, we have used definitions that closely approximated 1990 census categories: African American, Asian, American Indian, and Hispanic. So for 2000, the Asian and Native Hawaiian or other Pacific Islander groups have been combined. We computed segregation indexes using anyone designating himself or herself as a member of a particular racial group, e.g., Black or African American alone or in combination with another group (or groups).

The alternative was to use only individuals identifying with that group alone. Appendix A shows residential segregation indexes for 2000 calculated both ways.

We decided to use the "alone or in combination" method for two reasons. First, as described in Appendix A, using a different method had little impact on estimates of African American segregation and only a modest effect on those of Asians and Pacific Islanders and American Indians and Alaska Natives. Second, and perhaps more importantly, was that for some racial groups (particularly Native Hawaiians and other Pacific Islanders, but also American Indians and Alaska Natives) so many people chose more than one race that we were concerned that the analysis using only those who identify with one group alone would have excluded too many metro areas to provide reliable results.[1]

B. AREAS AND UNITS OF ANALYSIS

Residential segregation usually describes the distribution of different groups across units within a larger area. Thus, to measure residential segregation, we must define both the appropriate area and its component parts (its units of analysis). While residential seg-

regation can occur at any geographic level, we have chosen to focus on metropolitan areas as reasonable approximations of housing markets. The census-defined "place," which represents a town or city, is often too small. For example, some individuals in Washington, DC, need only move across the street to be in another jurisdiction, such as Prince George's County, Maryland. However, Consolidated Metropolitan Statistical Areas (CMSAs) seem too large; the New York CMSA stretches from Pennsylvania to Connecticut. We present estimates for all independent and primary metropolitan statistical areas (MSAs), referred to hereafter as metropolitan areas.[2]

The second geographic consideration — choosing an appropriate component part or unit of analysis — also presents alternatives. Independent estimates for racial characteristics are available for occupied households, census tabulation blocks, block groups, tracts, places, and counties. Both places and counties seem too large; movement from Park Avenue to Harlem in Manhattan, within the same place (New York City), or from Scarsdale to Yonkers, within the same county (Westchester County, New York) should have some measurable effect on segregation indexes.

---

[1] For historical analysis, Native Hawaiians and Other Pacific Islanders are combined with Asians in 2000 to calculate indexes comparable to the Asian and Pacific Islander population in 1980 and 1990. The residential segregation of Native Hawaiians and other Pacific Islanders in 2000 is analyzed in Chapter 4.

[2] OMB is introducing a substantially new concept for metropolitan areas to be defined on the basis of results of Census 2000 by June 30, 2003.

Occupied households are at the other end of the spectrum. Movement from one household to another usually occurs because of some life cycle event, and not to mitigate residential segregation.

That leaves blocks, block groups, and tracts. Blocks are created to ease data collection and can often have no residents, especially in commercial or industrial areas. Block groups are created by the Census Bureau as an intermediate geographic level to permit release of tabulated data that cannot be presented at the block level for confidentiality purposes. Arguments can be made that residential segregation indexes ought to be built up from the smallest geographic unit available — the block. Yet we believe it makes less sense to include the residents you may never see (on the opposite edge of a census block as blocks tend not to cross streets) and exclude the residents living across the street (in a different block). Going to larger aggregations of blocks, this problem is mitigated, although it never disappears as all geographies have boundaries.[3] Census tracts, which typically have between 2,500 and 8,000 people, are defined with local input, are intended to represent neighborhoods, and typically do not change much from census to census, except to subdivide. In addition, census tracts were often the unit of analysis chosen by other researchers. Consequently, we have chosen census tracts as our

unit of analysis.[4] We will examine the effects of choosing census block groups instead of tracts in future research.

C. RESIDENTIAL SEGREGATION MEASUREMENT

Residential segregation has been studied extensively with a variety of measures for many years (Duncan and Duncan, 1955; Taeuber and Taeuber, 1965; and Lieberson, 1980, 1981). Massey and Denton (1988) compiled, augmented, and compared these measures and used cluster analysis with 1980 census data from 60 metropolitan areas to identify five dimensions of residential segregation: evenness, exposure, concentration, centralization, and clustering. These five dimensions were further broken down into 20 measures of segregation, 19 of which we have calculated.[5]

- *Evenness* involves the differential distribution of the subject population.
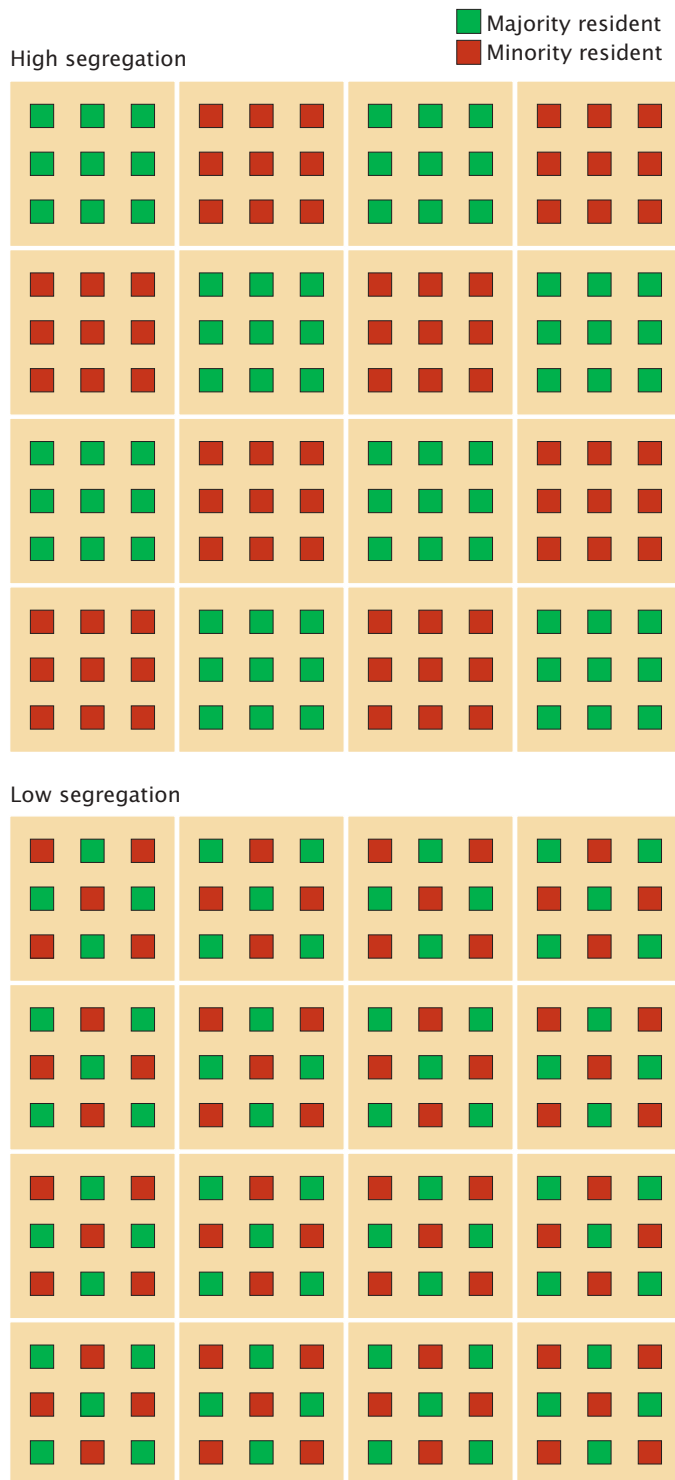
- *Exposure* measures potential contact.

- *Concentration* refers to the relative amount of physical space occupied.

- *Centralization* indicates the degree to which a group is located near the center of an urban area.

- *Clustering* measures the degree to which minority group members live disproportionately in contiguous areas.

Appendix B discusses all 19 measures proposed by Massey and Denton in detail. It also presents comparative analysis of the indexes within each dimension. Based on our assessment of the indexes, Massey and Denton's recommendations, and earlier research, we selected the indexes listed in Table 2-1 above to represent the five Massey-Denton dimensions.

The most widely used measure of evenness and the most-widely used measure of residential segregation, in general, is dissimilarity. Conceptually, dissimilarity, which ranges from 0 (complete integration) to 1 (complete segregation), measures the percentage of a group's population that would have to change residence for each neighborhood to have the same percent of that group as the metropolitan area overall.

Table 2-1.
Dimensions of Segregation and Indexes Used

Dimension of Segregation	Index Representing the Dimension
Evenness	Dissimilarity Index
Exposure	Isolation Index
Concentration	Delta Index
Centralization	Absolute Centralization Index
Clustering	Spatial Proximity Index

---

[3] One interesting future possibility is to tabulate data pooled across block faces, though this would take a great deal of work and must await better geographic information systems at the Census Bureau.

[4] We note that tract subdivision can increase measured residential segregation if it creates more homogeneous tracts.

[5] We omit an index which measures the proportion of the minority group residing in the central city of the metropolitan area. Massey and Denton (1988) note that this index, while quite simple to calculate, is a rather poor measure of segregation. We agree.

Figure 2-1a.
Distribution of Households Within One Hypothetical Metropolitan Area With High Segregation and One With Low Segregation:
Dissimilarity Index (Evenness)

■ Majority resident
■ Minority resident

High segregation

Low segregation

Note: Neighborhoods within the metropolitan areas are delineated by the white lines.

The exposure measure, the isolation index, describes "the extent to which minority members are exposed only to one another," (Massey and Denton, 1988, p. 288) and is computed as the minority-weighted average of the minority proportion in each area. It also varies from 0 to 1.

We chose delta as the measure of concentration. This index, which also varies from 0 to 1, measures the proportion of a group's population that would have to move across neighborhoods to achieve a uniform density across a metropolitan area. Massey and Denton's preferred concentration measure, relative concentration, does not conform well to theoretical constraints, having several calculated values below -1.

Absolute centralization examines only the distribution of the minority group around the metropolitan area center and varies between -1 and 1. Positive values indicate a tendency for group members to reside close to the center, while negative values indicate a tendency to live in outlying areas as compared with the reference group. A score of 0 means that a group has a uniform distribution throughout the metropolitan area.

Finally, the clustering measure used here, spatial proximity, basically measures the extent to which neighborhoods inhabited by minority members adjoin one another, or cluster, in space. Spatial proximity equals 1 if there is no differential clustering between minority and majority group members. It is greater than 1 when members of each group live nearer to one another than to members of the other group, and it is less than 1 in the rare case

that minority people lived nearer, on average to nonminority people than to members of their own group.

Figure 2-1(a-e) provides illustrations of what high and low segregation look like for all five measures; it shows how the measures capture different dimensions of segregation. Red boxes represent minority residents, while green ones represent majority residents. Each group of boxes represents a neighborhood, and each illustration represents a metropolitan area. Using the dissimilarity index (Figure 2-1a), a metropolitan area with high segregation has very homogeneous neighborhoods, though the location of those neighborhoods within the metropolitan area does not matter. Low segregation is characterized by an even distribution of minority group members across neighborhoods. In contrast, the isolation index, a measure of exposure, segregation (Figure 2-1b) is sensitive to the overall number of minority group members. Thus, the figure illustrating high segregation shows a metropolitan area with relatively few majority group members, and not evenly spread across tracts. Low segregation shows high levels of exposure to majority group members.

Figure 2-1b.
Distribution of Households Within One Hypothetical Metropolitan Area With High Segregation and One With Low Segregation:
Isolation Index (Exposure)
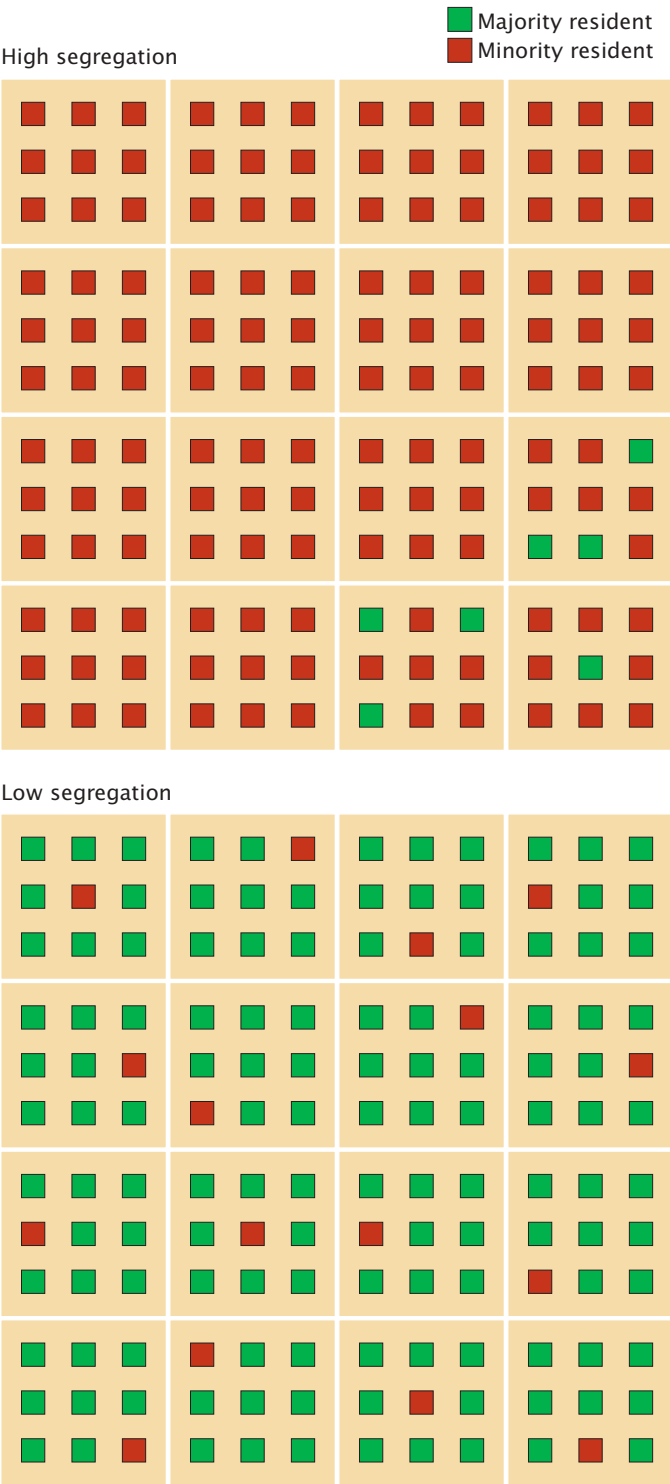
Note: Neighborhoods within the metropolitan areas are delineated by the white lines.

Figure 2-1c.
Distribution of Households Within One Hypothetical Metropolitan Area With High Segregation and One With Low Segregation:
Delta Index (Concentration)

■ Majority resident
■ Minority resident

High segregation

Low segregation

Note: Neighborhoods within the metropolitan areas are delineated by the white lines.
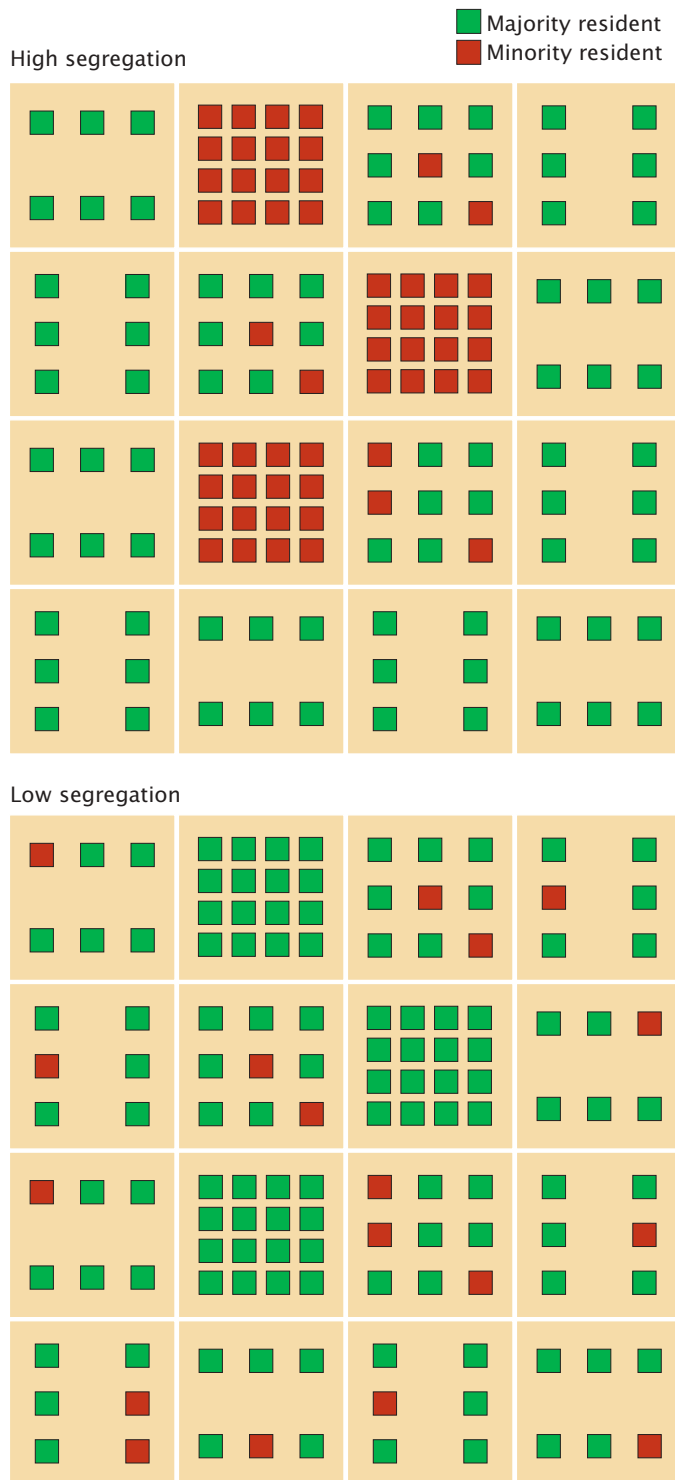
Metropolitan areas with high levels of concentration (Figure 2-1c), as measured by the delta index, are ones where minority members are densely packed in certain neighborhoods, while the low concentration illustration shows minority group members less densely packed in physical space than majority group members.

Figure 2-1d illustrates high centralization (the absolute centralization index), which measures the degree minority members are disproportionately in neighborhoods at the center of the metropolitan area, while low centralization indicates that minority group members are more toward the periphery of the metropolitan area.

Finally, clustering (Figure 2-1e), as measured by the spatial proximity index, is sensitive to the proximity of tracts to one another, regardless of how close to the metropolitan area center they are (centralization) or their density (concentration). So the illustration of high clustering shows that tracts with many minority group members are adjacent to each other, while the illustration of low clustering shows them further apart.

Because our choice in this report, to focus on five specific indexes, has subjective elements, the Internet materials accompanying this report have information on all 19 indexes, not just the five chosen. We note that the dissimilarity index is the one most often chosen by researchers calculating only one index.

Figure 2-1d.
Distribution of Households Within One Hypothetical Metropolitan Area With High Segregation and One With Low Segregation: Absolute Centralization Index (Centralization)
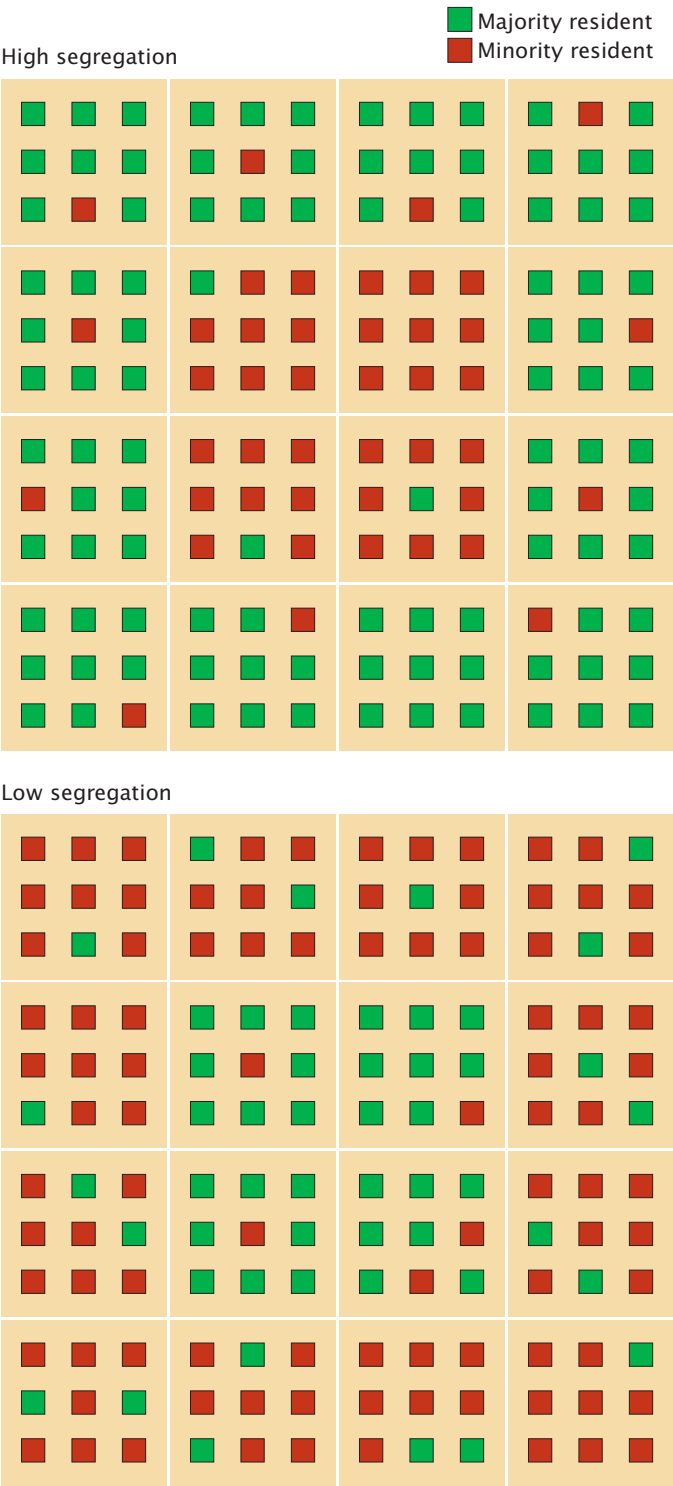
Note: Neighborhoods within the metropolitan areas are delineated by the white lines.

Figure 2-1e.
Distribution of Households Within One Hypothetical Metropolitan Area With High Segregation and One With Low Segregation:
Spatial Proximity Index (Clustering)

■ Majority resident
■ Minority resident

High segregation

Low segregation
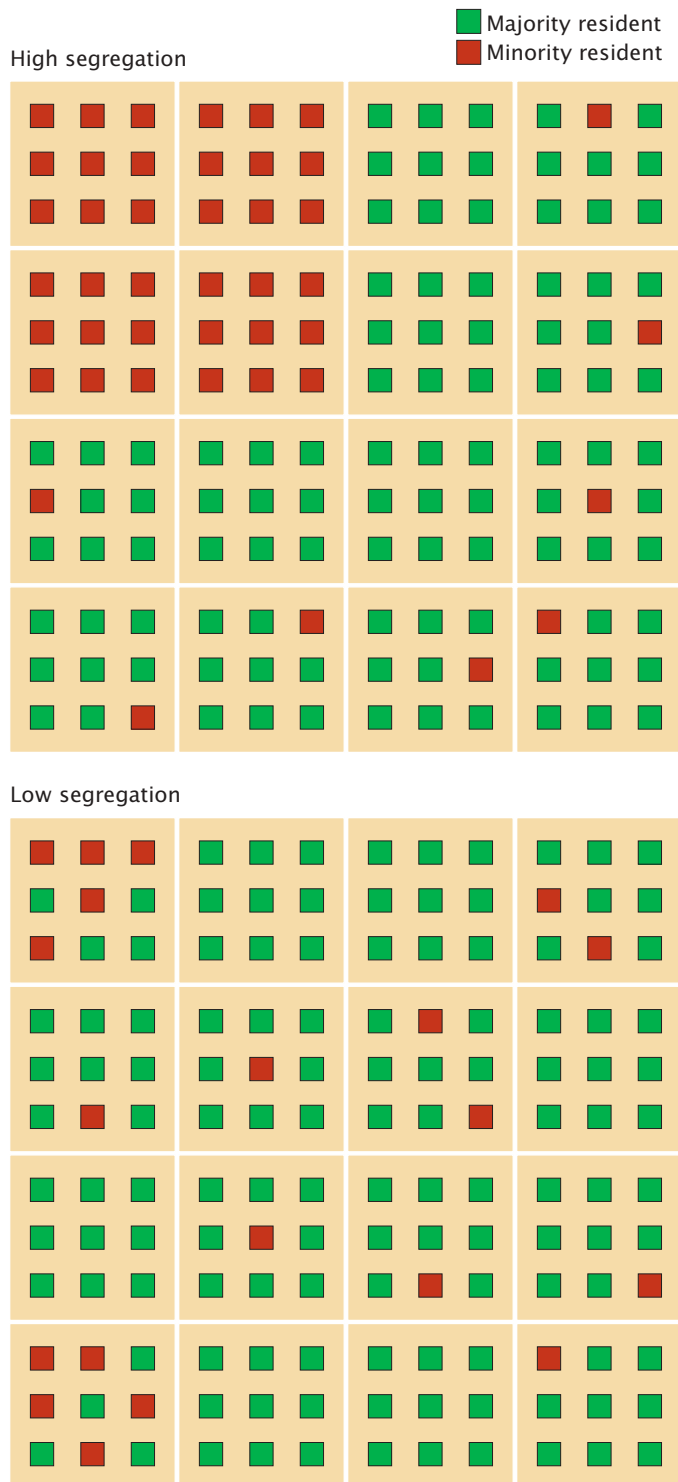
Note: Neighborhoods within the metropolitan areas are delineated by the white lines.

D. DATA

The data for this analysis were drawn from Census Bureau files (Census 1980, 1990, and 2000 Summary File 1) giving population counts for all racial groups and for Hispanics by census tract in all metropolitan areas. Data are presented for independent MSAs and Primary MSAs, not Consolidated MSAs. Town and city-based MSAs are used in New England. For 1980, 1990, and 2000 comparisons, the boundaries of metropolitan areas as defined on June 30, 1999 are used to ensure comparability.[6]

While this analysis uses constant metropolitan area boundaries, it does not use constant tract boundaries. The latter would require a considerable amount of mapping beyond the scope of this project. Tracts are sometimes added, split, or combined between censuses. Newly constructed tracts may tend to have greater racial or ethnic homogeneity than others, given that tracts are designed to represent relatively homogenous neighborhoods, and race may be one factor in their construction. The magnitude and effect of tract redefinition on computed segregation scores is not well understood.

Some estimates are presented at the aggregate summary level of "all U.S. metropolitan areas." Most estimates are for MSAs with a minority population of at least 20,000, or 3 percent of the 1980 total population.[7] We have imposed these restrictions because segregation indexes for metropolitan areas with small minority populations are less reliable than those with

---

[6] Counts may differ from official counts as tracts representing Crews of Vessels have been eliminated.
[7] MSAs must also have at least 10 census tracts.

larger ones. Random factors and geocoding errors are more likely to play a role in determining the settlement pattern of group members when fewer members are present, causing these indexes to contain greater variability. We note that Farley and Frey (1996) used these same cutoffs in their analysis. When averages across MSAs are presented, they are weighted by the minority group population in the MSA.

As a visual supplement to the segregation indexes, we created maps displaying the spatial distribution of racial and ethnic population groups in the most and least segregated large metropolitan areas for each group, as well as Los Angeles, which has an exceptionally diverse population. These maps are intended to convey a general impression of the racial and ethnic population distribution of these metropolitan areas, rather than showing exact numbers or locations. The map elements are scaled to a size that works well in the "typical" census tract. Consequently, in the largest and smallest census tracts, these maps do not always accurately display the true location and size of the underlying population.

When interpreting the maps, it is important to keep in mind that the underlying data consist of population counts in each census tract. The maps display dots of uniform size, each of which represents 200 or 400 people of a particular racial or ethnic group. Dots are placed arbitrarily within census tracts. This scheme works best in medium-sized census tracts. In large, sparsely populated census tracts, dots are not necessarily placed on the true population centers, and may be placed in nonresi-

Table 2-2.		
Estimated Net Percent Undercount From Demographic Analysis: 1990-2000		
	1990	2000
Total	1.65	0.12
Non-Black	1.08	-0.29
Black	5.52	2.78

Note: The non-Black category in that year refers more specifically to those who were not Black or non-Black Hispanic.

Source: Estimates are from demographic analysis (Robinson 2001, Table A).

dential areas, such as uninhabited mountains, airports, or industrial facilities. Other census tracts are too small and densely populated to fit enough dots, causing dots to overlay one another. When there is not enough space in a census tract to display dots representing both non-Hispanic Whites and the minority group, the minority group dots overlay and cover up dots representing non-Hispanic Whites.

E. STATISTICAL TESTING

Because the base data are from the decennial census, they have no sampling error and conventional tests of significance do not apply. Any criteria adopted to discern substantive, rather than statistical, differences in segregation scores is inevitably somewhat arbitrary. We designate substantively noteworthy index differences as those that are more than 1 percent of the range of the index estimates for metropolitan areas meeting the minimum size criteria. For example, in 2000, the dissimilarity index for American Indians and Alaska Natives ranged from 0.213 to 0.607, a range of 0.394. Thus, differences of 0.004 (1 percent of 0.394) are considered substantively notable for this index

for comparisons across MAs within this time period. For changes across time, the average of the 3 years' index range is used.[8]

Changes are shown in terms of percentage change in various tables. We present data in this way in order to make comparable statements across indexes whose ranges differ. It should be noted, however, that the small base of some index scores (i.e., those close to zero), may result in large percentage increases or decreases, even while the point change is small. Readers can refer to mean scores in the different years shown in various tables (or actual scores of different metropolitan areas as shown on the Internet) to compute change in different ways.

In some tables, we rank metropolitan areas according to their level of segregation, and we also average ranks across the five measures of segregation. We consider differences of less than 1 in the average rank to be basically tied. This cutoff, 1, was not derived based on any specific statistical procedure.

---

[8] There was little difference in this range among the years.

Apart from the issue of statistical testing, it should be noted that these data also have nonsampling error. Estimates of net undercoverage (underenumeration) of the total population are 1.65 percent for 1990, and 0.12 percent for 2000. This relatively low level of undercount masks differential undercount — a higher undercount of minority populations than non-minority ones. Table 2-2 shows estimated undercounts for the total population, Blacks, and non-Blacks in 1990 and 2000.

How this differential undercount affects residential segregation indexes is not known. If the people who were missed are distributed geographically like the people who were enumerated, then there may be little impact. Also, because of their complexity, segregation indexes are particularly subject to programming error. Appendix C discusses how the indexes calculated in this study compare with others.

F. INTERPRETING THE FINDINGS

We think it critically important to note that the values and ranks we report for metropolitan statistical areas on the several indexes can readily be misinterpreted as indicating that residential segregation is a more serious problem in some metropolitan areas, and a less serious problem in others. We strongly emphasize that the reported measures cannot necessarily sustain such inferences or interpretations. In particular, we do not speculate about how racial discrimination, free choices, or any of several other underlying processes (e.g., the growth or contraction of housing of varying costs relative to the growth or contraction of populations of varying incomes and stages of household formation; the relationship of such housing and population to jobs, schools, shopping and other amenities) might have contributed to the patterns observed. Similarly, the measures tell us nothing about consequences of an observed residential distribution (e.g., differential access to educational or job opportunities, a group's ability to maintain culturally distinctive institutions or practices) that might assist in identifying either problems or benefits associated with the pattern.

For these reasons, the measures reported here should be viewed as representing a starting point for research on contemporary patterns of residential segregation in the United States. To facilitate such work, as we noted above, the values for all 19 indexes for all metropolitan areas for each of the years and groups examined are available on the Internet.